Early Journal Content on JSTOR, Free to Anyone in the World

This article is one of nearly 500,000 scholarly works digitized and made freely available to everyone in the world by JSTOR.

Known as the Early Journal Content, this set of works include research articles, news, letters, and other writings published in more than 200 of the oldest leading academic journals. The works date from the mid-seventeenth to the early twentieth centuries.

We encourage people to read and share the Early Journal Content openly and to tell others that this resource exists. People may post this content online or redistribute in any way for non-commercial purposes.

Read more about Early Journal Content at http://about.jstor.org/participate-jstor/individuals/early-journal-content.

JSTOR is a digital library of academic journals, books, and primary source objects. JSTOR helps people discover, use, and build upon a wide range of content through a powerful research and teaching platform, and preserves this content for future generations. JSTOR is part of ITHAKA, a not-for-profit organization that also includes Ithaka S+R and Portico. For more information about JSTOR, please contact support@jstor.org.

THE VARIABILITY IN THE DISTRIBUTION OF WEALTH AND INCOME.

SUMMARY.

There is no generally accepted method of measuring the concentration of wealth, 416.—The methods of Lorenz, 417, Pareto, 419, and Watkins, 427, explained and analyzed.—The coefficient of variability, as used by statistical biologists; suggested as a good measure, 429.—Application of that measure to income statistics, 436, and to wealth statistics of Massachusetts, France, and the United Kingdom, 440.—Conclusion, and summary of the features of the different measures of variability, 447.

I.

BEFORE there can be any general agreement concerning the tendency in the distribution of wealth and income, two conditions must be fulfilled: first, adequate and reliable statistics; and, second, a scientific and generally understood method of measuring concentration. At the present time neither of these conditions is fulfilled. Wealth, income, and wage statistics are very meagre; and the articles of Messrs. Lorenz, Holmes, and Watkins in Nos. 70, 71, 72, and 81 of the Quarterly Publications of the American Statistical Association demonstrate that there is no generally accepted method of measuring concentration of wealth. It is the object of this paper to consider the various methods of measuring the inequality of the distribution of wealth that have been suggested, and to apply those methods to the available statistics.

Professor Lorenz has pointed out the mathematical inaccuracy of certain commonly used methods, and has suggested a graphic solution. Given the number of individuals holding estates of varying size, Professor Lorenz would arrange them in order, poorest to richest. He would then plot points on cross section paper whose abscissas are the percentages of the total wealth held and whose ordinates are the percentages that the holders of such wealth bear to the total number. A curve would then be drawn through such points. Equality of distribution would give a series of points in a straight line. The amount of inequality would be measured by the amount of deviation of the curve from the straight line. Dr. Watkins notes that "Mr. Lorenz uses hypothetical figures, and thus fails to perceive the limited practical value of the particular species of graphic method that he prefers."[1]

The writer has applied the method to the Massachusetts Probate Returns for the four periods 1829–31, 1859–61, 1879–81, and 1889–91, with the result shown in Chart I.[2] The curve for the period 1829–31 shows the least

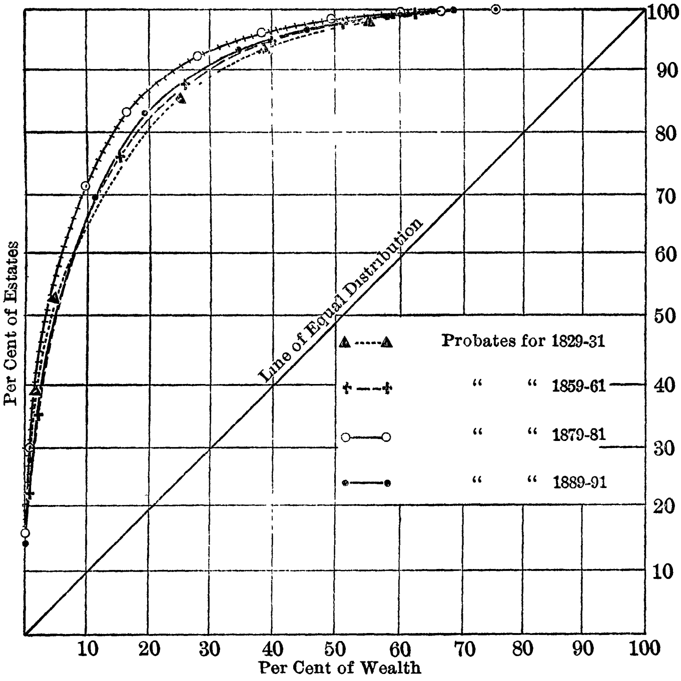


CHART I—APPLICATION OF THE METHOD OF LORENZ TO MASSACHUSETTS PROBATE RETURNS.

---

[1] Publications of American Statistical Association, No 81, p 33 (foot-note).

[2] Twenty-fifth Annual Report of the Bureau of Statistics of Labor, Massachusetts, 1894, p. 265.

bulge, except at the lower extremity, where it shows next to the greatest. Hence, according to this method of measurement, the least concentration exists among these estates. The periods 1859–61 and 1889–91 come next with about equal concentration, and the period 1879–81 has the highest concentration. All the curves are nearly coincident, for the lower 50 per cent. of the estates and for the most concentrated 40 per cent. of the wealth. The method is defective in not allowing comparison at these points. Again, the method requires that a comparison of the bulge of the curves be made by the eye, many times an inconclusive test.

In his turn Dr. Watkins suggests measuring concentration by the slant of certain curves.[1] Suppose that in any community we have given the number of estates as large as $1,000, the number of estates as large as $5,000, as large as $10,000, and so on, for two or more different periods. Dr. Watkins would obtain a curve for each period by plotting points whose abscissas are the logarithms of the numbers of estates, each of which estates is as large as the designated amount, and whose ordinates are the logarithms of the designated amounts. That is, the logarithm of the number of estates of $1,000 or over and the logarithm of $1,000 would be corresponding abscissa and ordinate. "The steeper of two curves," says Dr. Watkins, "is the one which expresses the greater concentration."[2] An illustration of the kind of curves obtained by application of Watkins's method is shown by Chart II.

The method used by Dr. Watkins is a variation of that developed by Professor Pareto. The exact connection between the two will appear upon consideration of the latter.

---

[1] Publications of American Statistical Association, No 81, p 34

[2] Ibid., p. 34.

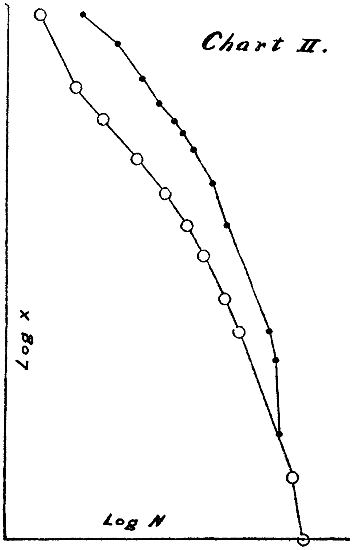


CHART II.—APPLICATION OF WATKINS'S METHOD TO BRITISH ESTATES OF £100 OR OVER IN 1901–02 TO 1905–06 (UPPER GRAPH WITH SOLID POINTS) AND FRENCH ESTATES OF 2,000 FRANCS OR OVER IN 1903–04 (LOWER GRAPH WITH CIRCLES).

II.

Professor Pareto, in his study of incomes, has observed that, if N represent the number of individuals having an income of x or greater, then the function $N = \frac{A}{x^{\alpha}}$ will fit the statistics closely, where A and α are constants to be determined from the given data. This law of the distribution of incomes is a purely empirical one, no *a priori* reason for it being given. Written in the logarithmic form, the equation becomes $\log N = \log A - \alpha \log x$, which, if $\log x$ be plotted as abscissa and $\log N$ as ordinate, is the equation of a straight line having a slope equal to $-\alpha$. Professor Pareto has suggested that α be used as a measure of the inequality of distribution.[1]

---

[1] Pareto's equation can be used only to represent the distribution of those incomes which decrease in frequency as the size of the income increases. Hence,

A numerical illustration may make clear the significance of the various terms in Pareto's equation. It is evident that A indicates the number of individuals earning one unit or more, for, if one be substituted for x in the equation, then, regardless of the value of α, $N = A$. Suppose that we have two communities of 1,000 individuals each, with incomes of one dollar or more distributed according to Pareto's law. Suppose, further, that for the first community $\alpha = 2$ and for the second community $\alpha = 3$. Inserting these values in $N = \frac{A}{x^{\alpha}}$, we have the following frequency tables (decimals omitted):—

COMMUNITY I.
($\alpha = 2$ and $N = 1{,}000$.)

Income.	*Number in Income Group.*
1 and less than 2	750
2 " " " 3	139
3 " " " 4	48
4 " " " 5	23
5 " " " 6	12
6 " " " 7	8
7 " " " 8	4
8 " " " 9	4
9 " " " 10	2
10 " " " 15	6
16 " " " 25	2
26 and upward	2
Total	1,000

COMMUNITY II.
($\alpha = 3$ and $N = 1{,}000$.)

Income.	*Number in Income Group.*
1 and less than 2	875
2 " " " 3	88
3 " " " 4	21
4 " " " 5	8
5 " " " 6	3
6 " " " 7	2
7 " " " 8	1
8 " " " 9	1
9 and upward	1
Total	1,000

it cannot be used to represent the usual distribution of wages, in which the number of wage-earners increases until a "modal" wage is obtained, and then decreases as wages increase.

It is to be noticed that the smaller value of α gives a wider range of distribution. This is true in general. The median, quartiles, and deciles can be computed easily by the use of the formula, by inserting the proper values for N and solving for x. The median of incomes over one unit is found by substituting $\frac{A}{2}$ for N; the quartiles, by substituting $\frac{3A}{4}$ and $\frac{A}{4}$; the deciles (lowest to highest), by substituting $\frac{9A}{10}, \frac{8A}{10}, \ldots\ldots\ldots \frac{A}{10}$. For the illustrations given we have the following results:—

	Community I.	Community II.
First quartile (Q_1)	1 155	1 101
Median	1 414	1.260
Second quartile (Q_2)	2.000	1.587

Substituting the quartile values in Bowley's measure of dispersion, $\frac{Q_1 - Q_2}{Q_1 + Q_2}$, we get 0.268 and 0.181 as the measures of dispersion for communities I. and II., respectively. This would indicate that the smaller value of α goes with the greater inequality of distribution.

That Professor Pareto's law apparently describes the distribution of incomes with a remarkable degree of accuracy is shown by its application to the statistics of income in Prussia for the years 1892 and 1902.[1] The points on Charts III. and IV. are located by plotting as abscissas the logarithms of the lower limits of classes of incomes from 900 marks upward, and as ordinates the logarithms of the numbers of incomes equal to or greater than such lower limits for the two years 1892 and 1902, respectively. Taking Pareto's equation, $\alpha \log x - \log A + \log N = O$, the values of α and $\log A$ were computed by the writer such that the sum of the squares of the distances of the

---

[1] Zeitschrift des Preussischen Statistischen Bureau, 1904, p 263.

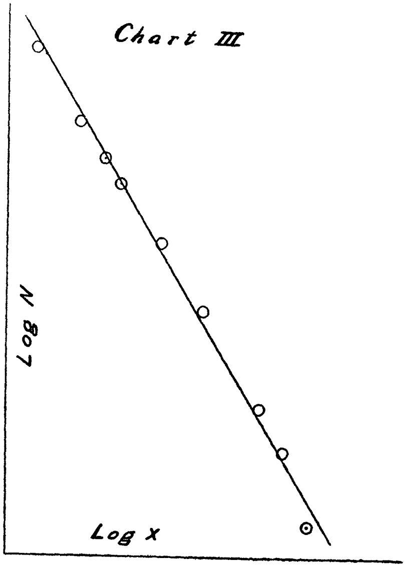


CHART III.—APPLICATION OF PARETO'S METHOD TO PRUSSIAN INCOMES OF 900 MARKS OR OVER IN 1892.

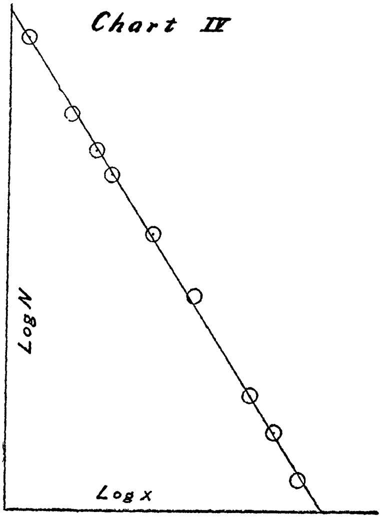


CHART IV —APPLICATION OF PARETO'S METHOD TO PRUSSIAN INCOMES OF 900 MARKS OR OVER IN 1902.

given points to the straight lines be a minimum. The respective lines about which the points are grouped are the lines best fitting the points according to the least square criterion. The equations of the lines are:—

$$1.656 \log x - 11.3395 + \log N = O \text{ for } 1892$$

and

$$1.570 \log x - 11.1767 + \log N = O \text{ for } 1902.$$

If inequality in the distribution of incomes increases as α decreases, then the inequality increased during the decennial period considered.[1]

That Pareto's law does not describe the distribution of estates as closely as it does the distribution of incomes is at once evident upon reference to Chart V. The solid points are located by the statistics of estates or portions of estates of £100 or more liable to estate duty in the United Kingdom for the five-year period 1901–02 to 1905–06. The circles are located by the statistics of French successions of 2,001 francs and over in 1903 and

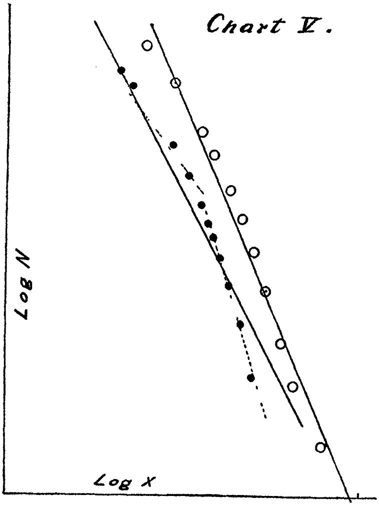


CHART V—APPLICATION OF PARETO'S METHOD TO THE FRENCH AND BRITISH STATISTICS USED IN CHART II

---

[1] See note on p 449

1904. In neither case is the "fit" nearly as good as it was for Prussian incomes. The dotted lines are the least square lines for the British estates of from £100 to £50,000, and from £75,000 upwards, respectively. Even tho the French and British data were fully comparable, a comparison of the values of α for the two series would not indicate the relative amount of concentration because of the decided variation of the points from a straight line. Thus the only reason that Pareto gives for the adoption of α as the measure of concentration—namely, the closeness with which the points group themselves about a straight line—fails for these statistics. A comparison of the values of α for the two sets of statistics thus means nothing.

Taking only those values of £10,000 or more from the British statistics, and of 250,000 francs or more from the French statistics, we have the configuration shown in Chart VI. The value of α for the British estates is

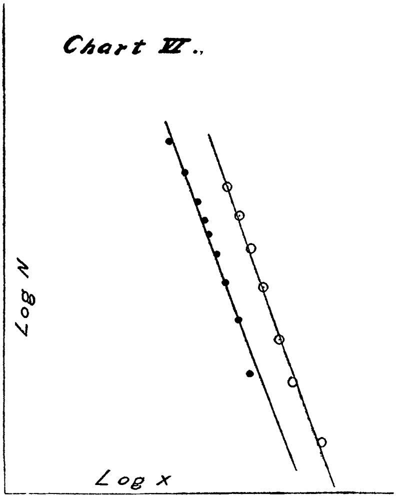


CHART VI —APPLICATION OF PARETO'S METHOD TO BRITISH ESTATES OF £10,000 OR OVER IN 1901–02 TO 1905–06 (GRAPH AT LEFT WITH SOLID POINTS) AND TO FRENCH ESTATES OF 250,000 FRANCS OR OVER IN 1903–04 (GRAPH AT RIGHT WITH CIRCLES).

1.407, and for the French estates is 1.422, indicating a slightly greater variability among British estates.

Before making any further application of Pareto's method, it is desirable to determine the degree of accuracy with which the equation $N = \frac{A}{x^{\alpha}}$ represents the statistics. The simple observation of the closeness of fit of the points in Charts III. and IV. to the straight lines there given is not a sufficient test of accuracy, since the co-ordinates of the points are logarithms, and an error in a logarithm gives a much larger error in the natural number. The relation between the errors in numbers and in the logarithms of those numbers is shown by the following table:—

1% error in	100	makes	0.216%	error in	log 100.
	1,000	"	0.144%	" "	log 1,000.
	10,000	"	0.108%	" "	log 10,000.
	100,000	"	0.086%	" "	log 100,000.
	1,000,000	"	0 072%	" "	log 1,000,000.
10% error in	100	makes	2.06%	error in	log 100.
	1,000	"	1.37%	" "	log 1,000.
	10,000	"	1.03%	" "	log 10,000.
	100,000	"	0.83%	" "	log 100,000.
	1,000,000	"	0.69%	" "	log 1,000,000.
25% error in	100	makes	4.84%	error in	log 100.
	1,000	"	3.23%	" "	log 1,000.
	10,000	"	2.42%	" "	log 10,000.
	100,000	"	1.94%	" "	log 100,000.
	1,000,000	"	1.61%	" "	log 1,000,000.

In order to test the accuracy of Pareto's method for those statistics for which the accuracy seems to be the greatest, the writer has used the Prussian income statistics. The procedure was as follows:—

1. Pareto's equations were computed by the least square method for Prussian incomes of (*a*) 3,000 marks or more in 1892, (*b*) 3,000 marks or more in 1902, and (*c*) 100,000 marks or more in 1902. The resulting equations are given in Table I. The data used in computing the equations are given in columns (2) and (5) of that table.

TABLE II.

Table showing accuracy with which Pareto's equation, $\alpha \log x - \log A + \log N = O$, gives the numbers of incomes in the various classes of incomes received in Prussia in 1892 and 1902.[1]

	1892.			1902.				
(1) Class.	(2) Actual Number Reported	(3) Number Estimated from Equation No 1 given below.	(4) Per Cent. of Error in Estimated Number in Column (3)	(5) Actual Number Reported.	(6 Number Estimated from Equation No 2 given below	(7) Per Cent. of Error in Estimated Number in Column (6).	(8) Number Estimated from Fquation No. 3 given below.	(9) Per Cent. of Error in Estimated Number in Column (8).
(Marks)			%			%		%
900 to 3,000	2,118,969	2,596,400	+22 6	3,309,696	2,791,500	—15 7	7,256,000	
3,000 to 6,000	204,544	275,900	+35 0	291,341	332,800	+14 4	743,300	
6,000 to 9,500	55,561	67,800	+22 0	77,636	87,310	+12 5	178,800	
9,500 to 30,500	46,092	50,482	+9 5	64,737	69,875	+8 1	130,180	
30,500 to 100,000	9,034	7,256	—19 7	13,205	11,395	—13 6	17,974	
100,000 to 500,000	1,555	1,083	—30.4	2,594	1,948	—24.9	2,569	—1 0
500,000 to 1,000,000	72	54	—25.0	108	114	+15.6	122	+13 0
1,000,000 to 2,000,000	27	17	—37 1	44	38	—13 6	38	—13.6
Over 2,000,000	4	8	+100 0	16	20	+25.0	17	+6.2
Total	2,435,858	2,999,000	+23.1	3,759,377	3,295,000	—12.3	8,329,000	—

Equation No 1 computed from incomes in 1892 over 3,000 marks is $1\ 668 \log x - 11\ 4037 + \log N = O$.
" No. 2 " " " " 1902 " " " " $1\ 560 \log x - 11.1263 + \log N = O$
" No. 3 " " " " 1902 " 100,000 " " $1\ 702 \log x - 11\ 9486 + \log N = O$

---

[1] Statistics from Zeitschrift des Preussischen Statistischen Bureau, 1904, p 255

2. In the equations thus computed the lower limits of the classes of incomes were inserted, and the numbers of incomes equal to or greater than such amounts were found.

3. The number of incomes in each class was then readily found by subtraction. These results are given in columns (3), (6), and (8) of Table I.

4. The per cent. of error in the number estimated to the actual number in each class was computed, and appears in columns (4), (7), and (9) of the table.

The percentage of error found is surprisingly large. Among the 1892 incomes of 3,000 marks or over, the percentage error in the various classes ranges from —37.1 to +35.0, except in the highest class, where the error reaches +100 per cent. Among the 1902 incomes of 3,000 marks or more the percentage error in the various classes ranges from —24.9 to +25.0. Even among the 1902 incomes of 100,000 or over the percentage error ranges from —13.6 to +13.0.

The preceding analysis shows that the accuracy of Pareto's first approximation is apparent and not actual. Pareto's method is not to be relied upon in comparisons where the difference in the amount of variability is not great.

III.

Chart II. gives the graphic representation of British estates of £100 or over and French estates of 2,000 francs or over, according to Dr. Watkins's method. The solid points are located by the British statistics, and the circles by the French statistics. Dr. Watkins's method is a slight variation of that of Pareto, obtained by simply interchanging ordinate and abscissa. The slope of the curve joining the points is taken as an index of variability. This test, as used by Dr. Watkins, is to be applied whether or not the points lie along a straight line. Thus, in order to measure variability, it is necessary to estimate the relative slopes throughout the curves, a method that may result in much difference of opinion. The test is uncertain and inconclusive. The method of Pareto, from which this method is derived, is meant to be applied only in those cases where the points arrange themselves along a straight line. In such cases the values of α can be computed and compared. On the other hand, Dr. Watkins applies his method, however the points be situated, and offers no numerical measure of variability. That this leads to difficulties and inconsistencies will be shown later.

The objections to Dr. Watkins's method may be summarized as follows:—

1. The exact meaning of the test is not evident. As the author himself says, "Just what the degree of difference is between two series of numbers so compared is not obvious."[1]

2. The comparison of slopes of the curves has to be made entirely by the eye, which is a very unreliable scientific instrument. No numerical measure is offered. If, in comparing two curves, the second is steeper at one point than the first, and not so steep at another point, there is no way of telling the net result.

3. If the average size of the estates within a class change or if the largest estates should increase indefinitely in size, the slope of the curve would not change. That the average size of the estates in any one class does change considerably from year to year is brought out by the British statistics given in Table XI. The average amount of the estates over £1,000,000 varies from £1,000,000 in 1838 to £3,022,000 in the last period. This decided change does not affect the slope of Watkins's curve, as it would if the slope were a true measure of concentration.

---

[1] Publications American Statistical Association, No. 81, p 33

4. The effect of the smaller estates in the comparison of slopes does not seem to be given its true importance. The lower part of the curve is left out in several cases by Dr. Watkins in making his comparisons. Thus in the comparison of Massachusetts probate statistics for four periods, 51 per cent., 35 per cent., 83 per cent., and 27 per cent., respectively, of the number of estates are omitted in constructing the curves.

All of the methods which have been considered in the preceding analysis will be applied later in connection with the discussion of still another method of measuring distribution which will now be developed.

IV.

Mathematicians and astronomers have long had occasion to use an index of variability. In measuring any physical quality a large number of times by any instrument, however accurate, the astronomer finds that the resulting measurements are not all alike, but show variations or errors.

> The errors here referred to are the small accidental errors which necessarily affect all human observations, owing to the imperfection of the senses and of the instruments employed, as well as to any unfavorable conditions under which they are used. Such are as likely to be errors in excess as errors in defect, and hence in the long run tend to counteract each other.[1]

If the observer be untrained, the instrument crude, and the conditions unfavorable, then the variation among the measurements taken will be great. If the observer be trained, the instrument refined, and the conditions favorable, then the variation among the measurements taken will be correspondingly small. The necessity of finding a numerical index of variation among measurements is evident. Astronomers have adopted as their index of variability the square root of the mean square of the deviation of the individual measurements from the arithmetic average of the measurements. This quantity has been termed the standard deviation, usually represented by the Greek letter σ.[1] The *precision* of any set of measurements varies as the reciprocal of the standard deviation.

---

[1] Greene, Spherical and Practical Astronomy, p. 115.

The statistical biologists have also found it necessary to measure variation. Evolution, according to the Darwinian theory, is dependent upon variability among individuals. "Here in variety, the various deviations from the racial type, we find the material for selection."[2] The standard deviation used by the astronomers has been adopted as the measure of variability for measurable biological characteristics. As Karl Pearson says,

> other measures of variability have been devised, some of which are occasionally useful, but for theoretical and practical reasons the standard deviation may be considered the best. It is not hard to find, and it occurs and recurs in all sorts of investigations.[3]

---

[1] Let σ = standard deviation,
n = number of measurements.
X = any measurement,
$X_1 \dots X_n$ = the individual measurements,
m = the arithmetic mean $= \frac{X_1 + \quad + X_n}{n}$
x = deviation of any item from the arithmetic mean $= X - m$.
$x_1 \dots x_n$ = deviation of the measurements $X_1 \quad X_n$ from the arithmetic mean,

Then $\sigma = \sqrt{\frac{x_1^2 + x_2^2 + \dots + x_n^2}{n}}$,

or $\sigma = \sqrt{\frac{\Sigma x^2}{n}}$ (Σ stands for "the sum of such terms as")

If each of the individual measurements occurs several times, say X_1 occurring f_1 times, X_2 occurring f_2 times, and so on, where $f_1, f_2, \dots f_n$ stand for the frequencies, the formula becomes

$$\sigma = \sqrt{\frac{f_1x_1^2 + f_2x_2^2 \quad + f_nx^2_n}{n}}, \text{ or } \sigma = \sqrt{\frac{\Sigma fx^2}{n}}$$

where $n = f_1 + f_2 + \dots + f_n$.

[2] Karl Pearson, The Grammar of Science (2d edition), p 384

[3] Ibid., p. 387.

The statistical problem before the economist in determining upon a measure of the inequality in the distribution of wealth is identical with that of the biologist in determining upon a measure of the inequality in the distribution of any physical characteristic. In order to determine the variability of the stature of men in a certain community, the investigator would need a statistical table giving the number of individuals of the varying heights. Such a table is called a frequency table. From this frequency table the standard deviation could be found easily. In order to determine the variability of the income or wealth of men of any community, such statistics would be substituted for those of stature.

That the standard deviation actually does measure inequality of distribution can be understood fully only by computing it for statistics, hypothetical and actual, and noticing the change following a change in the distribution of the measurements. Let us consider the following set of hypothetical statistics·—

Daily Wage.	1	2.	3.	4	5.	6.	7.	8.	9.	10.
The number receiving such wage in:										
Period I.	120	30	20	13	7	4	2	2	1	1
Period II.	100	50	15	20	15					

Total number of wage-earners in each period = 200.
Average daily wage in each period = $2.

Our object is to obtain a numerical measure of the inequality of distribution of the wages in each of the two periods. The entire range of wages in Period I. is from $1 to $10, and in Period II. the range is from $1 to $5. Within each range occurs the variation. However, the range alone does not adequately measure the variation. The inequality in distribution depends upon the number of persons receiving the indicated wages as well as upon the range. The total daily amount of wages paid to the 200 wage-earners of each period is $400. The average wage is \$2 per day. It is evident that, if every one of the 200 wage-earners should receive \$2 per day, there would be no inequality in distribution. The measure of inequality under such circumstances *ought* therefore to be zero. Inequality begins as soon as some receive more and others receive less than the arithmetic average wage. The first step in figuring the standard deviation is to obtain the deviation of each wage from the arithmetic average, which deviations are given in column 3 of the illustrations. Column 4 of the illustrations contains the product of each deviation and the number of wage-earners having such deviation. Column 5 is obtained by taking the product of corresponding items in columns 3 and 4. The square root of the total of column 5 divided by the number of wage-earners (200) gives the standard deviation. The standard deviations in the two illustrations, 1.664 and 1.285, are indices of the variability of the two groups of statistics.

ILLUSTRATIONS.

I.

1. Wage (X).	2 Frequency (f)	3 Deviation (x)	4 fx	5 fx^2
1	120	—1	—120	120
2	30	0	0	0
3	20	+1	+20	20
4	13	2	26	52
5	7	3	21	63
6	4	4	16	64
7	2	5	10	50
8	2	6	12	72
9	1	7	7	49
10	1	8	8	64
Total	200	—	—	554

$$\text{Standard deviation} = \sqrt{\frac{\Sigma fx^2}{n}}.$$

$$= \sqrt{\frac{554}{200}}.$$

$$= 1\ 664$$

$$\text{Coefficient of variation} = \frac{\sigma}{n} .100 = 83.2.$$

II

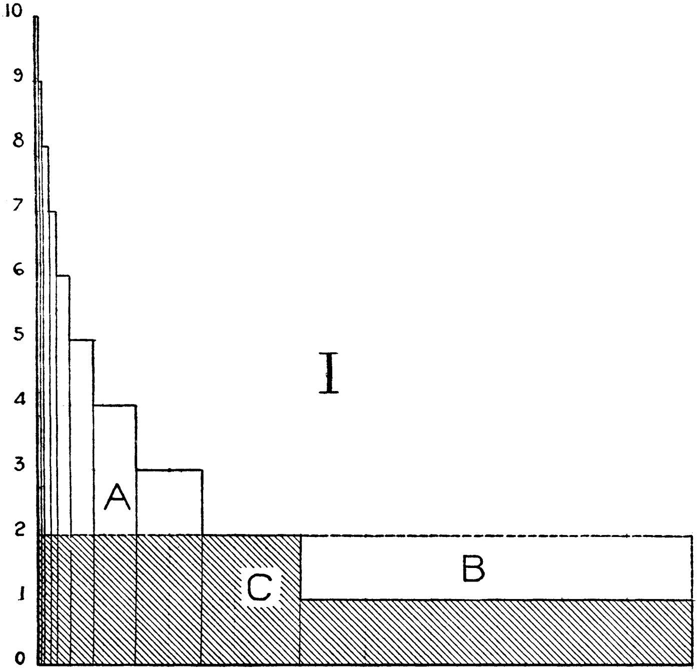
1	100	—1	—100	100
2	50	0	0	0
3	15	+1	+15	15
4	20	2	40	80
5	15	3	45	135
Total	200	—	—	330

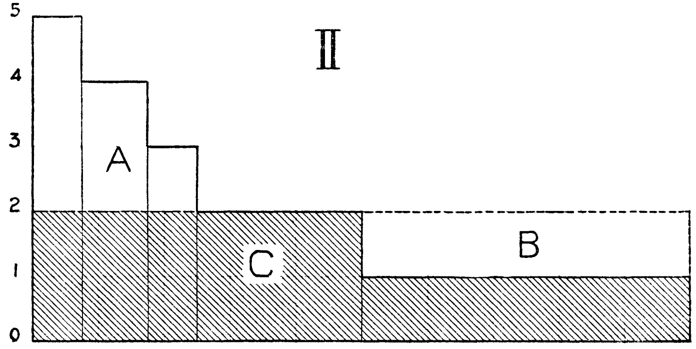
Coefficient of variation = 64 2

However, if the average daily wage had not been the same ($2) in the two cases, a comparison of the standard deviations would not have shown the relative variability. In order to obtain measures which are comparable, no matter what the arithmetic average may be, the standard deviation should be divided by the arithmetic average. This ratio (the standard deviation to the arithmetic average) multiplied by 100 gives the *coefficient of variability.* The arithmetic average and the coefficient of variability are complementary measures: the first indicates the type, while the second indicates the amount of variation from that type.

The manner in which a change in the statistics affects the coefficient of variability may be understood better by referring to the following graphic representation of the figures given in the illustrations. The heights of the rectangles appearing in Figures I. and II. correspond to daily wages received, and the widths are proportional to the numbers receiving such daily wages. The rectangles are arranged from left to right in order of their altitudes. The area of each rectangle corresponds to the total amount of daily wages received by some group of wage-earners. The area enclosed by the axes and the broken line corresponds to the total daily wages of the 200 workmen. A horizontal dotted line is drawn at a height corresponding to the average wage. The areas *A* and *B* are equal respectively to the sum of the positive and of the negative deviations. Area *A* = Area *B*. The standard deviation, and therefore the coefficient of variability, does not only depend upon the *sizes* of the areas *A* and *B*, but also on their *shapes*.

Consider Figure I. Suppose that the total amount of wages and the number of wage-earners remain fixed, but that the distribution of the wages change, what would be the effect on the figure? If the wage-earners receiv-


10
9
8
7
6
5
4
3
2
1
0
I
A
C
B


5
4
3
2
1
0
II
A
C
B

ing \$3 per day be reduced to \$2 per day, and the wages of the \$4 and \$5 men be increased one dollar each, the coefficient of variability for the new distribution would be 90. If twenty of the wage-earners receiving \$1 per day be reduced to nothing per day, and the wages of the \$4 and \$5 men be increased one dollar each, the coefficient of variation would be 95.5. In both cases the change means a greater concentration of wages. In the first case there was a shifting of the wages among those receiving more than the average, but that shifting was away from rather than towards equality of distribution. When, as in the second case, the increase of wages of those already receiving more than the average is at the expense of those receiving less, the tendency towards inequality is obviously doubly great. As the concavity of the jagged line (tops of rectangles) becomes greater, the inequality in distribution must be greater.

That a uniform percentage increase or decrease in all wages will not affect the coefficient of variability is easily demonstrated. If each wage be multiplied by a, the arithmetic average (m) will be multiplied by a, and therefore each deviation (being $aX—am$) would be multiplied by a. The standard deviation becomes $\sqrt{\frac{\Sigma fx^2a^2}{n}}$, or $a\sqrt{\frac{\Sigma fx^2}{n}}$, and the coefficient of variability becomes $\frac{a\sigma}{am}.100$, or $\frac{\sigma}{m}.100$.

Likewise it may be demonstrated that, if the *number* in each group of wage-earners be increased by a given percentage, the coefficient of variability will be unchanged. Suppose each frequency to be multiplied by b, the total number of wage-earners will be bn. The standard deviation will be $\sqrt{\frac{\Sigma bfx^2}{bn}}$, or $\sqrt{\frac{b\Sigma fx^2}{bn}}$. Since the arithmetic average and standard deviation are unchanged, the coefficient of variability will be unchanged.

V.

Professor Henry L. Moore has applied the above-described method of measuring variability to the most complete wage statistics that have been collected in the United States,—those in the "Employees and Wages" volume of the Twelfth Census.[1] The following table gives the number of employees, average weekly wage, and coefficient of variability for the thirty industries selected:—

	1890.	1900.
Number of laborers	104,923	160,055
Average weekly wage	$11.57	$11.52
Coefficient of variability	45.9	43.5

Both the average weekly wage and the coefficient of variability decreased during the decennial period. The decreased variability in the group of laborers considered may be due to the labor unions in their effort to equalize wages, to the further introduction of machinery, and the decrease of highly paid skilled laborers, or other reasons. Whatever may be decided upon as the *cause* of the decrease of variability, it is to be noted that Professor Moore has shown that the variability of the wages of this large group of laborers considered was less in 1900 than in 1890. This is a fact that must be considered in any complete study of the position of the wage-earner.

It would be extremely illuminating if we could obtain the tendency as to variability of wages over a long period, say from thirty to fifty years. As Professor Moore points out,[1] there are numerous important questions that can be answered only by such an investigation. The only series of wage statistics in the United States that covers an extensive period, and which is in such form that the coefficient of variability can be computed, is that of the Aldrich Report. These statistics, however, have been found to be far from satisfactory. The very narrow scope of the investigation is shown by Table II.

---

[1] Political Science Quarterly, March, 1907, p 61.

TABLE II.

Table compiled from the Aldrich Report, giving number of male employees in each industry in July of 1860, 1870, 1880, and 1890.

Industry.	1860.	1870.	1880.	1890.
1. Agricultural Implements	8	40	43	51
2. Ale, Beer, and Porter	41	36	79	55
3. Books and Newspapers	48	84	171	153
4. Building Trades	392	727	747	842
5. Carriages and Wagons	14	24	40	40
6. City Public Works	2,464	1,641	559	716
7. Cotton Goods	480	540	535	592
8. Dry Goods	0	5	6	9
9. Ginghams	219	312	535	627
10. Groceries	3	5	6	6
11. Illuminating Gas	475	507	385	503
12. Leather	46	88	112	83
13. Lumber	14	22	9	8
14. Metals and Metallic Goods	851	1,207	1,643	1,813
15. Paper	27	15	16	13
16. Railroads	126	305	321	516
17. Sidewalks	38	21	16	17
18. Spice	7	7	18	26
19. Stone	443	641	697	690
20. White Lead	22	20	22	18
21. Woollen Goods	149	237	339	433
Total	5,867	6,484	6,299	7,201

Altho the data of the Aldrich Report are insufficient and results drawn from them are not even to be taken as showing the conditions in the States or industries covered, yet they may be used to further *illustrate* the methods of measuring variability. The result of tabulating the male wage-earners according to weekly wages in groups, each one dollar wide, is given in Table III. The arithmetic average and coefficent of variability for each year are given in Table IV. For the group of wage-earners considered, both the average wage and the variability have increased. However, as noted above, no sure conclusion can be drawn as to the distribution of wages even in the industries considered. In 1860 42 per cent. of the total number of employees came under the caption

---

[1] Ibid. "What answer, based upon actual data, can the economist give to these questions: Does skilled labor hold its own in the shifting proportions of grades of labor in the various industries? Are the chances of promotion within an industry increasing or decreasing? Is the status of the laborer confined within more narrow or less narrow limits?"

TABLE III.

Table compiled from the Aldrich Report, giving the number and per cent. of wage-earners receiving the indicated weekly wage in July of 1860, 1870, 1880, and 1890.

Rate per Week.	1860		1870.		1880.		1890.	
	Number	Per Cent	Number	Per Cent.	Number.	Per Cent.	Number	Per Cent.
Less than $2	84	1 43	—	—	4	06	—	—
$2 and less than $3	57	97	3	05	76	1 21	18	.25
3 “ “ “ 4	97	1 65	109	1 68	115	1 82	120	1 67
4 “ “ “ 5	142	2 42	73	1 13	109	1 73	98	1 36
5 “ “ “ 6	342	5 83	29	45	45	71	57	79
6 “ “ “ 7	2,663	45 40	115	1 77	319	5 06	215	2 95
7 “ “ “ 8	523	8 91	193	2 98	1,235	19 60	592	8 23
8 “ “ “ 9	317	5 40	99	1.53	313	4 97	597	8 30
9 “ “ “ 10	433	7 37	453	6 98	699	11 09	994	13.85
10 “ “ “ 11	249	4 25	665	10 28	697	11 07	729	10 12
11 “ “ “ 12	73	1 24	67	1 03	135	2 14	150	2 08
12 “ “ “ 13	756	12 89	631	9 73	700	11 10	758	10 53
13 “ “ “ 14	66	1 13	1,466	22.60	633	10 04	857	11 90
14 “ “ “ 15	4	07	104	1 57	117	1 86	254	3 53
15 “ “ “ 16	27	.41	613	9 45	386	6 14	484	6.72
16 “ “ “ 17	1	.02	452	6 98	207	3 29	305	4 24
17 “ “ “ 18	—	—	26	40	30	42	29	.40
18 “ “ “ 19	18	31	414	6 38	291	4 62	223	3 10
19 “ “ “ 20	12	20	40	62	18	29	108	1.50
20 “ “ “ 21	1	02	43	66	15	24	39	54
21 “ “ “ 22	—	—	328	5 06	27	44	284	3 95
22 “ “ “ 23	—	—	31	48	12	19	58	81
23 “ “ “ 24	1	02	87	1 34	72	1 14	116	1 61
24 “ “ “ 25	—	—	136	2 10	25	40	77	1 07
25 “ “ “ 26	—	—	253	3 90	4	06	6	08
26 “ “ “ 27	—	—	—	—	—	—	1	01
27 “ “ “ 28	—	—	9	14	6	10	5	07
28 “ “ “ 29	—	—	2	03	—	—	4	06
30 “ “ “ 31	—	—	38	59	5	08	16	.22
31 “ “ “ 32	—	—	—	—	—	—	2	03
32 “ “ “ 33	—	—	—	—	1	02	—	—
33 “ “ “ 34	—	—	1	02	—	—	—	—
35 “ “ “ 36	—	—	1	02	—	—	1	01
38 “ “ “ 39	1	02	2	03	2	03	2	.03
47 “ “ “ 48	—	—	—	—	—	—	1	01
95 “ “ “ 96	—	—	1	02	1	02	—	—
143 “ “ “ 144	—	—	—	—	—	—	1	01
Total	5,867	99 96	6,484	100 00	6,299	99 87	7,201	99.93

"City Public Works," while in 1890 only 10 per cent. were so classified. If such employees be omitted in 1860, the arithmetic average wage becomes $7.99, and the coefficient of variability becomes 37.9. The fact that no sure conclusion can be drawn as to the variability of wages from the statistics does not disprove the adequacy of the method. The tabulation necessary to apply the method shows the inadequacy of the statistics.

TABLE IV.

Arithmetic average and coefficient of variation of wages compiled from the Aldrich Report for years 1860, 1870, 1880, and 1890.

	1860.	1870.	1880.	1890.
Arithmetic average	$7.84	$14.37[1]	$11.05	$12.37
Coefficient of variability	34.7	35.8	39.3	39.9

[1] Greenback standard.

Both the coefficient of variability and α of Pareto's equation have been computed for various groupings of Prussian incomes for the years 1892 and 1902. The figures are given in Table V. The same conclusion as to concentration is indicated in each case by each method of measurement; namely, greater concentration in 1902.

TABLE V.—PRUSSIAN INCOMES.

	α of Pareto's Equation.		Coefficient of Variability.	
	1892.	1902.	1892.	1902.
All incomes	—	—	564	675
Of 900 m. or more	1.656	1.570	443	542
Of 3,000 m. or more	1.668	1.560	316	383
Of 100,000 m. or more . . .	1.945	1.702	128	144
Of less than 100,000 m. and more than 900 m. . . .	1.538	1.549	—	—

It is to be noticed that the coefficient of variability decreases as the range of incomes is made less. On the other hand, α of Pareto's equation does not necessarily decrease. It is not legitimate to compare the concentration of different ranges of income by means of the coefficient of variability, but it is legitimate to compare αs for different ranges, if α be considered a good measure of variability. By the very process of computing α, the law of distribution over the entire range is assumed.

VI.

The best available statistical tables in regard to the distribution of wealth in the United States, England, Germany, and France, have been brought together by Dr. Watkins in the article on the Statistical Evidence of Concentration of Wealth, already referred to.[1] It is not the object here to make a critical study of the statistics there given to determine their merits as being representative of the actual conditions in the various countries, but to take the statistics as being the best material available and to determine whether or not they show a tendency towards or away from greater variability. The question at hand is, "Given the statistics, what do they indicate?"

Dr. Watkins has compared the French statistics given in Table VIII. and the United Kingdom statistics shown in Table IX. The logarithmic curves for the United Kingdom estates of £100 and upwards and for the French successions of 2,000 francs and upwards are given in Chart II.

Upon examination of the curves Dr. Watkins concludes, "The figures for 1903 and 1904 subjoined indicate considerably less present concentration of riches in France than in the United Kingdom."[2] . . . The coefficient of variability for the United Kingdom estates over £100 is 635, while the coefficient of variability for the French successions over 10,000 francs is 709, for successions over 2,000 francs is 1,080, and for all successions is 1,620. Therefore, the conclusions from the comparison of the coefficients of variability is at variance to the conclusion from a comparison of the slopes of the logarithmic curves. Which is correct?

---

[1] Publications American Statistical Association, No. 81.

[2] Ibid , p. 47 Dr Watkins says in a foot-note that "the difference is great enough to warrant an inference, even though the statistics are not exactly comparable." . . .

In order to get a third measure of concentration to compare with the two leading to opposite conclusions, the writer has adopted a slight variation of the test of Professor Lorenz. Charts VII. and VIII. show the upper right-hand corner and the lower left-hand corner of the rectangle of accumulated percentages of estates and wealth

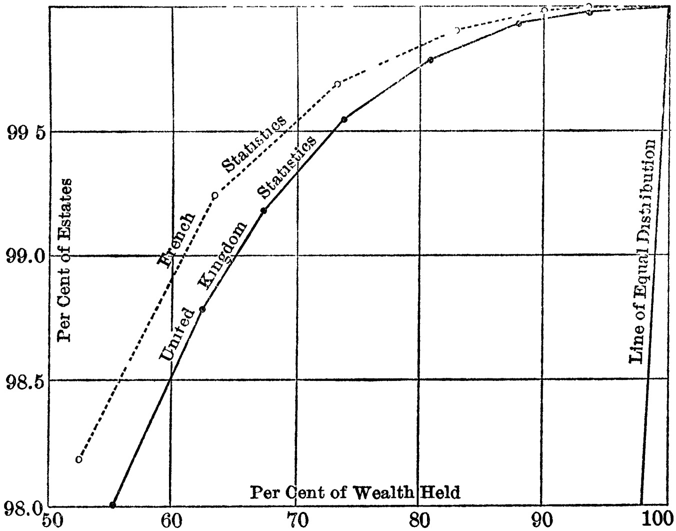


CHART VII.—APPLICATION OF METHOD OF LORENZ TO BRITISH ESTATES OF £100 OR MORE IN 1901–02 TO 1905–06 AND FRENCH ESTATES OF 2,000 FRANCS OR MORE IN 1903–04. (UPPER RIGHT-HAND PORTION OF RECTANGLE OF ACCUMULATED PERCENTAGES.)

held according to the French successions of 2,000 francs and upwards and the United Kingdom estates of £100 and upwards. Table VI. gives the percentages. The vertical and horizontal scales have been chosen so that the extremities of the graphs could be compared. The greater bulge of the French graph indicates greater concentration among the French successions than among

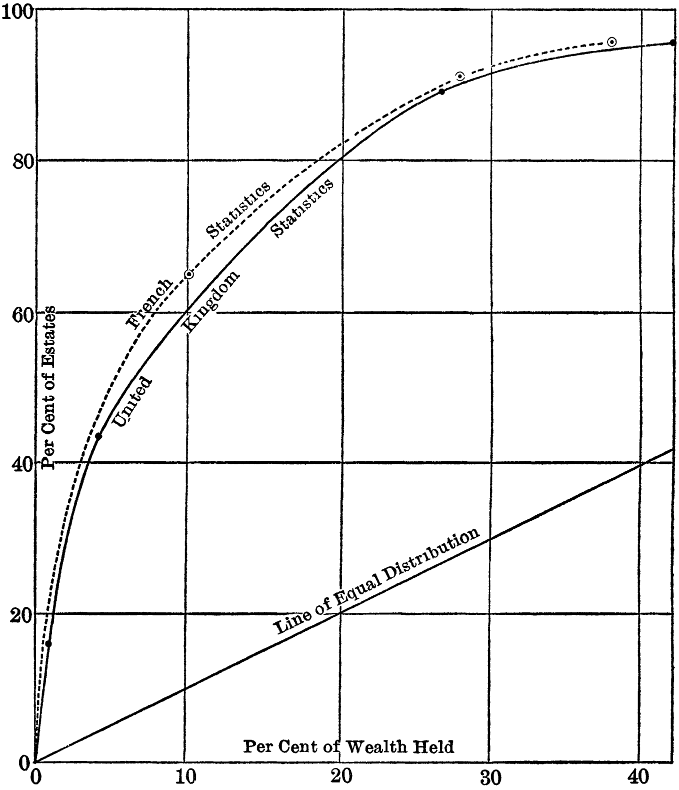


CHART VIII —APPLICATION OF METHOD OF LORENZ TO SAME STATISTICS AS WERE USED IN CHART VII. (LOWER LEFT-HAND PORTION OF RECTANGLE OF ACCUMULATED PERCENTAGES.)

United Kingdom estates. This test leads one to the same conclusion as the comparison of the coefficients of variability. It is safe to conclude, therefore, that the comparison of the slopes of the logarithmic curves is not a reliable test of concentration.

The coefficients of variability for the statistics given in Tables VIII. to XII. have been computed as follows:—

(1) The number of estates and average amount of estates in each class were determined.

(2) The true arithmetic average size of all of the estates for each period was computed.

(3) Deviations of the class average from the general average were found, and the squares multiplied by the number of estates in each class.

(4) From the data thus obtained the coefficients of variability were found by inserting in the formula as given on page 430.

A summary of the coefficients of variability is given in Table VII.

TABLE VI

United Kingdom, 1901–02 to 1905–06.		France, 1903 to 1904	
Percentage of Estates Richest to Poorest. (Over £100)	Percentage of Wealth held by Such Estates	Percentage of Estates[1] Richest to Poorest. (Over 2,000 francs)	Percentage of Wealth[2] held by Such Estates.
%	%	%	%
.0157	6 38	00692	6.28
.0713	12 22	02263	9.96
2095	19 24	.09873	17 19
.4494	26 18	.30773	26 75
.8142	32 73	75673	36.79
1.2077	37 63	1 72273	47 71
1 9957	44 86		
4 6	58.1	4 6	61 8
10 9	73 4	8 9	71 7
56 4	95 9	35 3	89 9
84 0	99 0		

[1] Number of estates, 318,154.

[2] Total amount, 9,868,784 thousand francs.

TABLE VII

COEFFICIENTS OF VARIABILITY.

1 *Massachusetts Probates.*

	1829–31.	1859–61	1879–81	1889–91
All Estates	510	445	527	426
Estates of $5,000 or over	242	241	311	252

2 *United Kingdom Estates Liable to Estate Duty*

	1838	1884–85 to 1888–09	1901–02 to 1905–06.
Estates of £100 or over	—	—	635
Estates of £1,000 or over	293	360	485
Estates of £10,000 or over	—	—	270

3 *French Successions.*

	1903–04
All Successions	1,620
Successions over 2,000 francs	1,080
Successions over 10,000 francs	709
Successions over 250,000 francs	279

TABLE VIII.

SUCCESSIONS IN FRANCE ACCORDING TO NET VALUE, 1903 AND 1904: CLASS, AVERAGE AMOUNT, AND NUMBER[1]

Class.	Average Amount	Number
(Francs)	(Francs)	
From 1 to 500	263	241,097
" 501 to 2,000	1,280	208,382
" 2,001 to 10,000	4,880	205,957
" 10,001 to 50,000	21,300	83,889
" 50,001 to 100,000	69,900	13,955
" 100,001 to 250,000	156,000	8,872
" 250,001 to 500,000	351,000	3,073
" 500,001 to 1,000,000	693,000	1,430
" 1,000,001 to 2,000,000	1,421,000	664
" 2,000,001 to 5 mil	2,950,000	242
" 5 to 10 mil	7,260,000	50
" 10 to 50 mil	17,700,000	18
Upwards of 50 mil	75,300,000	4
Total	—	767,633
Average succession	13,284	—

---

[1] Computed from table given in Watkins's "Statistical Evidence of the Concentration of Wealth," p. 48.

TABLE IX

CLASS, AVERAGE AMOUNT, AND NUMBER OF ESTATES OR PORTIONS OF ESTATES LIABLE TO ESTATE DUTY IN THE UNITED KINGDOM FOR THE FIVE-YEAR PERIOD 1901–02 TO 1905–06[1]

Class.	Average Amount	Number.
£100 and less than £500	£458	28,551
500 “ “ “ 1,000	837	49,136
1,000 “ “ “ 10,000	3,680	81,140
10,000 “ “ “ 25,000	18,200	11,221
25,000 “ “ “ 50,000	38,900	4,500
50,000 “ “ “ 75,000	68,300	1,403
75,000 “ “ “ 100,000	93,000	699
100,000 “ “ “ 150,000	132,000	649
150,000 “ “ “ 250,000	216,000	427
250,000 “ “ “ 500,000	378,000	246
500,000 “ “ “ 1,000,000	783,000	99
1,000,000 and upwards	3,022,000	28
Total	—	178,099
Average	£7,440	—

[1] United Kingdom Statistical Abstract, 53d number, p. 45.

TABLE X.

PROBATES IN MASSACHUSETTS CLASS, AVERAGE AMOUNT, AND NUMBER[1]

Class.	1829–31		1859–61		1879–81		1889–91.	
	Average Amount	Number	Average Amount	Number	Average Amount	Number	Average Amount	Number.
Under $500	$185	1,431	$234	1,485	$245	1,822	$247	2,217
$500 to 1,000	733	463	726	960	722	1,451	731	1,738
1,000 to 5,000	2,370	1,274	2,400	2,827	2,460	4,588	2,430	6,197
5,000 to 10,000	6,790	295	6,910	797	6,980	1,421	6,980	1,969
10,000 to 25,000	15,470	157	15,360	507	15,420	1,023	15,440	1,498
25,000 to 50,000	35,200	42	34,900	168	34,000	410	34,600	480
50,000 to 100,000	73,200	25	69,500	92	70,500	218	69,800	265
100,000 to 200,000	134,200	6	129,000	52	139,700	111	136,700	134
200,000 to 300,000	—	0	237,500	18	240,000	37	249,000	45
300,000 to 400,000	320,000	2	341,000	7	346,500	22	339,000	22
400,000 to 500,000	415,400	1	474,500	3	438,300	10	453,000	13
500,000 and over	634,000	2	848,000	6	1,145,000	29	795,000	30
Total	—	3,698	—	6,922	—	11,142	—	14,608
Average probate,	$3,920	—	$7,690	—	$12,330	—	$10,650	—

[1] Computed from tables given in an article on Statistical Evidence of the Concentration of Wealth, by G. P Watkins, Quarterly Publications of the American Statistical Association, No. 81, pp 30–31

TABLE XI.

CLASS, AVERAGE AMOUNT, AND NUMBER OF ESTATES OVER £1,000 EACH SUBJECT TO PROBATE DUTY IN THE UNITED KINGDOM IN 1838, 1884–85 TO 1888–89 AND 1901–02 TO 1905–06 [1]

Class.	(1) 1838		(2) 1884–85 to 1888–89		(3) 1901–02 to 1905–06	
	Average Amount	Number	Average Amount	Number	Average Amount	Number
£1,000 to £10,000	£3,040	6,006	£3,200	55,767	£3,680	81,140
10,000 to 50,000	19,380	901	21,150	10,145	24,100	15,721
50,000 to 100,000	68,400	83	70,000	1,251	76,500	2,102
100,000 to 500,000	163,200	35	188,500	737	205,000	1,322
500,000 to 1,000,000	600,000	2	654,000	43	782,000	99
1,000,000 and over	1,000,000	1	1,734,000	12	3,022,000	28
Total	—	7,028	—	67,955	—	100,412
Average	£7,026	—	£9,834	—	£12,663	

[1] Computed from tables given in Watkins's "Statistical Evidence of the Concentration of Wealth," pp 40–41

(1) Amount of personal property subject to probate and administration duties in the United Kingdom, 1838

(2) Estates assessed to probate duty in the United Kingdom during the five-year period 1884–85 to 1888–89 (exclusive of property assessed to corporation duty)

(3) Estates or portions of estates liable to estate duty in the United Kingdom for the five-year period 1901–02 to 1905–06

TABLE XII

CLASS, AVERAGE AMOUNT, AND NUMBER OF INCOMES IN PRUSSIA IN 1892 AND 1902 [1]

Class	1892		1902	
	Average Amount	Number	Average Amount	Number
(Marks)				
Under 900	458	8,726,215	465	9,053,608
900 to 3,000	1,372	2,118,969	1,350	3,309,696
3,000 " 6,000	4,070	204,544	4,050	291,341
6,000 " 9,500	7,440	55,561	7,410	77,636
9,500 " 30,500	15,500	46,092	15,450	64,737
30,500 " 100,000	49,900	9,034	49,600	13,205
100,000 " 500,000	178,000	1,555	187,000	2,594
500,000 " 1,000,000	653,000	72	666,000	108
1,000,000 to 2,000,000	1,370,000	27	1,455,000	44
Over 2,000,000	5,000,000	4	4,310,000	16
Total	—	11,162,073	—	12,812,985
Average income	870	—	997	—

[1] Computed from table given in Watkins's "Statistical Evidence of the Concentration of Wealth," p. 52.

The following conclusions may be drawn from a study of the coefficients of variability:—

(1) Probates in Massachusetts showed the greatest concentration in 1879–81 and the least concentration in 1889–91.

(2) The concentration among dutiable estates in the United Kingdom has been increasing during the period from 1838 to 1906.

(3) The indications are that there is less concentration in Massachusetts than in the United Kingdom. (Compare the variability among Massachusetts probates of $5,000 and over with that of United Kingdom estates of £1,000 and over.)

(4) The indications are that there is greater concentration among French than among United Kingdom estates.

(5) The tendency among Prussian incomes is toward greater concentration.

(6) Average estates and average incomes have increased in every case examined. The increase in that average is due to a disproportionate increase of the larger estates, except in the case of Massachusetts.

Any thorough examination of the available statistics concerning wealth distribution forces upon one an appreciation of the inadequacy of the statistical data and especially of the inadequacy of statistics for the United States.

VII.

The method that has been used by the writer in considering the various measures of the inequality in the distribution of wealth is entirely empirical. Professor H. L. Moore has recently pointed out with approval the purely utilitarian or pragmatic character of the argument of statisticians.[1] He quotes from Jevons's *Theory of Political Economy* as follows: "The deductive science of economics must be verified and rendered useful by the purely *empirical* science of statistics."[2] No excuse, therefore, is offered, or need be offered, for the use of empiricism.

In conclusion, the various methods developed may be considered briefly. All of the methods meet the test of relativity. A uniform percentage change in the amount of wealth held by each individual will not affect the measure of inequality. Lorenz's method certainly does measure concentration, but has the defects, first, of making comparison at the extremities of the curves extremely difficult, if not impossible, and, second, of not providing a numerical measure of concentration. Pareto's method provides us with a numerical measure which seems to be a good measure in some cases, but analysis has shown that the accuracy with which the law describes the distribution of wealth is merely apparent, and not actual. Watkins's method allows too much of the subjective to enter. The measurement by the eye of relative average curvatures is not satisfactory.

The coefficient of variability is recommended as the most satisfactory measure of variability because:—

1. All of the statistics of the frequency table enter into the determination of this *numerical* measure.

2. The larger and the smaller estates are automatically weighted by the use of the squares of the deviations. No other more satisfactory method of weighting has been suggested.

3. The deviations are taken from the arithmetic average, which is perfectly definite and easy to compute. In the writer's mind the arithmetic average best repre-

---

[1] "Statistical Complement of Pure Economics," Quarterly Journal of Economics, November, 1908, p. 16 *et seq.*

[2] Ibid , p 5. The Italics are mine

sents the "plane of distribution." If the median were used, some theory of distribution would have to be assumed in order to estimate it from the frequency table.

4. It is not difficult of computation. With the help of computing devices, such as the slide rule and tables of squares, the work is by no means excessive.

Objection may be offered to the coefficient of variability because tolerably complete statistics are necessary in order that it may be used. Answer is to be given that only when we do have complete statistics is *any* comparison of variability to be depended upon.

WARREN M. PERSONS.

DARTMOUTH COLLEGE.

NOTE TO p. 423.

The medians and quartiles, and Bowley's quartile measure of dispersion for Prussian incomes of M. 900 or over, computed for the equations

$$1.656 \log x - 11.3395 + \log N = 0 \quad (1892)$$
$$1.570 \log x - 11.1767 + \log N = 0 \quad (1902)$$

are given in the following table:—

	1892.	1902.
Median	M. 1488.	M. 1327.
Q_1	1165.	1025.
Q_2	2262.	2063.
$\frac{Q_2-Q_1}{Q_2+Q_1}$	32.0%	33.6%